

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 1 035 733 A2

(12)

## EUROPEAN PATENT APPLICATION

(43) Date of publication:  
13.09.2000 Bulletin 2000/37

(51) Int. Cl.<sup>7</sup>: H04N 5/45

(21) Application number: 00104761.2

(22) Date of filing: 04.03.2000

(84) Designated Contracting States:  
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE  
Designated Extension States:  
AL LT LV MK RO SI

(30) Priority: 11.03.1999 FI 990537

(71) Applicant:  
Nokia Multimedia Terminals Oy  
00045 Helsinki (FI)

(72) Inventor: Altonen, Janne  
20100 Turku (FI)

(74) Representative:  
Stendel, Klaus et al  
Nokia GmbH,  
Patentabteilung,  
Postfach 101823  
44718 Bochum (DE)

### (54) Method and arrangement for combining video pictures

(57) The invention relates to the insetting of a moving picture in a moving main picture when the picture signals are in an encoded digital format. The main idea is that the picture signals are combined prior to decoding. The frames of the picture to be inset are scaled down by reducing the number of macro blocks in them in such a manner that the picture whole is retained. The macro blocks of a certain area of the main picture are replaced by the macro blocks of the reduced picture and the combined video signal is decoded. The video sig-

nals (ES1, ES2) to be combined may be picked up from different sources or extracted from a transport stream where they are in packets. The system according to the invention requires only a single decoder (210), which considerably reduces the amount of computation required by the combination of the pictures. The advantage is emphasized if there are several pictures to be combined.

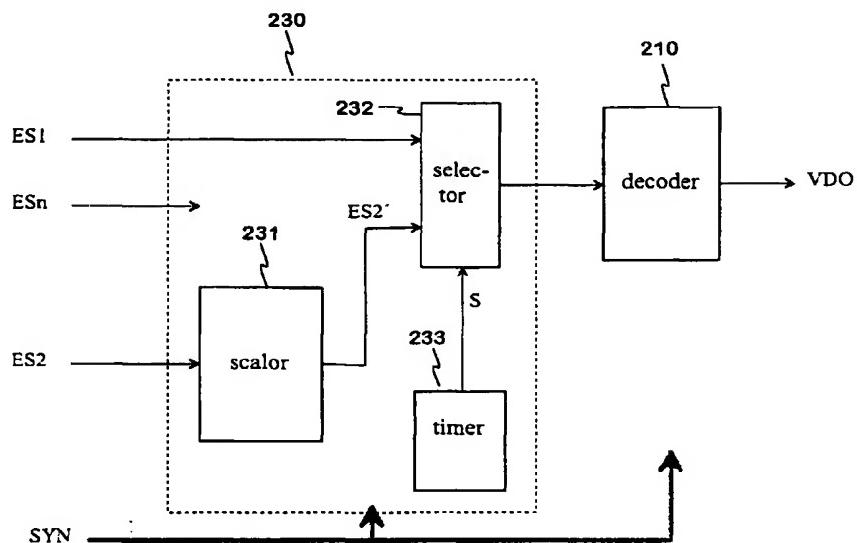


Fig. 2

## Description

**[0001]** The invention relates to a method for adding a moving picture on top of a larger moving picture. The method is applicable to cases in which the picture signals are compressed digital signals. The invention also relates to an arrangement for adding a moving picture on top of a larger moving picture.

**[0002]** The picture-in-picture (PIP) feature is a widely used technique at the transmitting end when making a television program, for example. Within the main picture a considerably smaller picture is temporarily inset which displays, say, a simultaneous event that is likely to interest the viewer. The addition of a secondary picture to the main picture may also occur at the receiving end, controlled by the user of the receiver. One or more smaller pictures may e.g. display programs running on other channels while the main picture is being viewed without interruption. The present invention relates particularly to the use of the PIP feature at the receiving end.

**[0003]** When compressing a digitized video signal it is customary to divide an individual picture frame into blocks which typically comprise 8x8 picture elements, or pixels. A square portion of a frame, comprised of four blocks, is called a macro block. Compression, i.e. reduction of the number of bits, is realized using intra-frame coding and inter-frame coding. The former includes e.g. predictive coding utilizing positional redundancy or the use of discrete cosine transform (DCT) concentrating signal energy of a picture block. The numbers produced by the transform are quantized in a manner that reduces the quantity of bits and ordered in a sequence where the occurrence of strings of consecutive zeroes is statistically high. These strings of zeroes are represented by a number indicating the quantity of the zeroes (this is called run length coding, RLC). Other numbers are encoded such that a frequently occurring number is represented by fewer bits than a number that occurs less frequently (variable length coding, VLC). Inter-frame coding includes predictive inter-frame coding, motion estimation comparing the contents of macro blocks at different positions, utilizing temporal redundancy and temporal interpolation of reference frames in order to produce the code for the frames between them. By combining various coding methods it is possible to reduce the number of bits transmitted down to a hundredth part of the original without substantially compromising the quality of the picture. At the receiving end of the video signal a video decoder performs the reverse operations. On the transmission path, several video signals may travel packet switched in the same transport stream (TS) so that the receiver first has to extract an individual video signal from it.

**[0004]** From the prior art it is known a PIP method wherein two encoded video signals are separately decoded and combined after the decoding. Prior to combining, one of the video pictures is reduced in size.

This can be done by selecting e.g. every fourth block in both the horizontal and vertical dimension of the video signal or by producing by means of interpolation a new macro block from each 4x5 macro block group. At a desired location of the normal-sized picture the macro blocks are then replaced by the macro blocks of the reduced picture. In this description and in the claims such a reduced picture is called a "mini-picture". The prefix "mini" means that the inset picture does not cover the whole main picture. Fig. 1 shows in the form of functional block diagram such a system according to the prior art. The system comprises decoders 110 and 120 as well as a PIP unit 130. A video signal ES1 (so-called "elementary stream") is brought to decoder 110 and video signal ES2 to decoder 120. Signals ES1 and ES2 are encoded e.g. according to the MPEG2 (Motion Picture Experts Group) standard. Decoder 110 outputs video signal VD1 and decoder 120 video signal VD2. The PIP unit 130 comprises a scaling unit 131, selector 132 and timing unit 133. Signal VD1 is directed straight to the selector. Signal VD2 is directed to the scaling unit 131 the output signal VD2' of which is conducted to the selector. The output signal VDO of the selector 132 is either signal VD1 or signal VD2' depending on the status of the selection signal S output by the timing unit 133. Always when the picture-generating system enters the area intended for the mini-picture, signal S goes into a state that conducts signal VD2' to the output of selector 132. At other times, signal S in a state that conducts signal VD1 to the output of selector 132. The functional blocks shown in Fig. 1 are realized partly in software and partly in hardware.

**[0005]** A disadvantage of the method described above is that it requires a double decoding operation. So, when using a signal processor, a double decoding capacity is required of it, which results in considerable extra costs. Another disadvantage is that in practice, for the reason stated above, only one mini-picture may be inset in the main picture.

**[0006]** An object of the invention is to eliminate the above-described disadvantages associated with the prior art. The picture-in-picture operation according to the invention is characterized by what is expressed in the independent claims. Preferred embodiments of the invention are presented in the dependent claims.

**[0007]** The main idea of the invention is that the video signals are combined prior to the decoding. The encoded macro blocks of the area of the main picture intended for the mini-picture are replaced by macro blocks from another picture. These are obtained e.g. by taking macro blocks at regular intervals in such a manner that their total number equals the number of macro blocks that corresponds to the mini-picture area. A single mini-picture macro block may also be produced by assembling it from selected blocks of several original macro blocks or by interpolating a plurality of original macro blocks into a single macro block. The combined signal is then decoded.

**[0008]** An advantage of the invention is that a receiver only needs one decoder. In video processing, decoding is the part that requires the most computing. Another advantage of the invention is that several mini-pictures can be inset in the main picture with only a minimum added computing. A further advantage of the invention is that the decoder is detached from the rest of the receiver so that the capacity of the transmission system between the decoder and receiver may be dimensioned according to the band of one channel only. Yet another advantage of the invention is that if the video signals to be combined are brought to the receiver in the same transport stream, it is only required one complete demultiplexer which extracts from the transport stream the packets relating to all auxiliary activities as well.

**[0009]** The invention is below described in detail. Reference will be made to the accompanying drawing in which

- Fig. 1 shows a block diagram illustrating the principle of an arrangement according to the prior art,
- Fig. 2 shows a block diagram illustrating the principle of the arrangement according to the invention,
- Fig. 3 shows an example of a PIP image on a screen,
- Fig. 4 shows a flow diagram illustrating the operation of the structure according to Fig. 2, and
- Fig. 5 shows in the form of block diagram an example of a system applying the invention.

**[0010]** Fig. 1 was already discussed in conjunction with the description of the prior art.

**[0011]** Fig. 2 shows in the form of functional block diagram a PIP implementation according to the invention. It comprises a PIP unit 230 and decoder 210. The PIP unit comprises a scaling unit 231, selector 232 and timing unit 233. Video signals ES1 and ES2 comprised of codes of horizontally adjacent macro blocks of a picture are brought to the system. Signal ES1 is conducted direct to the selector 232. Signal ES2 is conducted to the scaling unit 231 in which the picture is reduced using a known method. The output signal ES2' of the scaling unit is conducted to the selector 232. The output signal CS of the selector is either signal ES1 or signal ES2', depending on the status of the selection signal S output by the timing unit 233. Always when signal ES1 contains the code of a macro block meant for the area intended originally for the mini-picture on the screen, signal S is in state that directs signal ES2' to the output of selector 232. At other times signal S is in state that directs signal ES1 to the output of selector 232. Signal S is generated for the video signals ES1 and ES2 from temporally bound synchronization signals SYN, which are used to synchronize the operation of other units, too. Signal CS is conducted to decoder 210 which outputs a complete digital video signal VDO. Compared to

the structure of Fig. 1 this structure has one decoder less, which means an almost fifty percent drop in the need for computing capacity since decoding requires much more computation than the PIP function. Applying the principle according to Fig. 2 it is possible to inset several mini-pictures in the main picture. For each mini-picture it is needed a separate scaling unit in block 230 and extensions to selector 232 and timing unit 233. One common decoder is still enough, which emphasizes the advantage over the prior art.

**[0012]** Fig. 3 shows an example of an image produced by signal VDO of Fig. 1 or 2. It has a main picture 31 and an inset mini-picture PIP.

**[0013]** Fig. 4a shows in the form of flow diagram an example of the operation of the scaling unit 231. A complete frame comprises C macro blocks, or macros, horizontally and R macros vertically. In the example, every  $j^{\text{th}}$  macro is selected both horizontally and vertically in the frame of the picture to be reduced, the selected macros constituting signal ES2' of Fig. 2. In step 401 program reception is started. In step 402, the processing of an individual frame is started. In step 403, the values of variables r, i, c and j needed in the processing are initialized. Variable r is the number of a macro row in the complete picture, and variable c is the number of a macro column in the complete frame. Variable i is a row number, counting from the row that was last used to select macros for a mini-picture. Variable j is a column number, counting from the column that was last used to select a macro for a mini-picture. In step 404, row numbers r and i are incremented. In step 405 it is checked whether the last row of the frame was already processed. If so, the processing of the next frame is begun (step 402). If the frame is unfinished, it is checked according to step 406 whether the row being processed is a row on which macros are to be selected. If not, the column number c is incremented, step 407. In step 408 it is checked whether the macro row has come to an end. If not, the process moves on to the next macro on the row, step 409, and repeats steps 407 and 408. If the macro row has come to an end, the column number c is reset in step 410 and the process continues at step 404. If the row is a row on which macros are to be selected, the row number i is reset (step 411) and column numbers c and j are incremented (step 412). In step 413 it is checked whether the macro row has come to an end. If not, the process moves on to the next macro, step 414. In step 415 it is checked whether the column processed is a column on which macros are to be selected. If not, the process continues at step 412. If it is, the macro is saved according to step 416. At the same time column number j is reset. The process then continues at step 412.

**[0014]** Fig. 4b shows the area of a mini-picture. It comprises vertically  $R'$  macro block areas and horizontally  $C'$  macro block areas. Corresponding to the markings of Fig. 4a, the number of rows  $R'$  equals the ratio  $R/l$  rounded off to the nearest smaller integer, and the

number of columns C' equals the ratio C/I rounded off to the nearest smaller integer. The mini-picture stalls vertically from row r1 of the complete frame and horizontally from column c1 of the complete frame.

**[0015]** Fig. 4c shows in the form of flow diagram an example of the operation of the timing unit 233. A logic element or program corresponding to the diagram has at its disposal the values of variables r and c produced by the operation according to Fig. 4a. In step 421 the value of the row variable r and the value of the column variable c are read. In step 422 it is checked whether the current position in the complete frame is on a row that belongs to the mini-picture area. If not, selection signal S is set to zero (step 424), which state corresponds in Fig. 2 to the selection of signal ES1 by selector 232. If the current row falls within the mini-picture area, it is checked according to step 423 whether the current position in the complete frame is in a column that belongs to the mini-picture area. If not, the process moves on to step 424. If the current column falls within the mini-picture area, selection signal S is set to one (step 425), which state corresponds in Fig. 2 to the selection of signal ES2' by selector 232. The selling of signal S is realized in a synchronized manner between two consecutive macro block times. After the selling, the values of variables r and c are read again. Operation corresponding to Fig. 4c may also be realized such that the comparison corresponding to step 422 is made after the value of variable r has been incremented, and the comparison corresponding to step 423 is made after the value of variable c has been incremented.

**[0016]** In the operation according to Figs. 4a, 4b and 4c the position of the mini-picture to be inset in the main picture is determined by means of parameters r1, c1. The size of the mini-picture is determined by parameter I. The width-to-height ratio of the mini-picture is thus the same as that of the main picture. The ratio can be made freely selectable if an additional parameter is introduced in the program, determining how many columns or rows of the mini-picture frame produced as described above are included in the final mini-picture.

**[0017]** Fig. 5 shows an example of a system in which pictures are combined in accordance with the invention. The system comprises a front end 551 which receives a radio-frequency signal and outputs a base-band digital transport stream signal TS. Signal TS comprises consecutive packets comprised of a header and transport data proper. The header comprises, among other things, the packet identity data (PID). The transport stream may include packets containing the code of several different video signals and, in addition, packets associated with various auxiliary activities of the receiver. Signal TS is conducted to a TS multiplexer 541. This is a complete demultiplexer, which means it monitors a relatively large amount of PID numbers and extracts from the transport stream the respective packets. The demultiplexer 541 sends the code ES1 of the selected video signal to a selector 545 and the data

contents of the other extracted packets to a host processor 560 in the system. Signal TS is also conducted to a second TS demultiplexer 542. This is a relatively simple demultiplexer which extracts from the transport stream only the packets associated with a particular video signal. Of these the demultiplexer 542 produces video signal ES2 and feeds it to a PIP unit 530, to a scaling unit in it. In this example, video signal ES3 from a video disc drive 552 is also brought to selector 545. Signal ES3 is encoded in the same manner as signals ES1 and ES2. Selector 545 outputs the main picture signal ESs, which is either ES1 or ES3, depending on a control issued by the host processor 560. Signal ESs is conducted to the PIP unit 530, which corresponds to the PIP unit 230 in Fig. 2. Signal ES3 is also conducted direct to the PIP unit 530, to a scaling unit in it. Unit 530 reduces the number of macro blocks of signal ES2, signal ES3 or both, depending on a control issued by the processor 560. Furthermore, unit 530 substitutes macro blocks of the scaled-down pictures for a certain portion of the macro blocks of signal ESs corresponding to the main picture. Thus in this example it is possible to inset one or two pictures in the main picture. The output signal MP of unit 530 is conducted to a common decoder 510 which outputs a complete digital video signal VDO. It is used to generate the analog or digital signals controlling the display. The system also comprises a unit 561 to receive selection and control data from the user. The host processor 560 is connected with the other units via a bus 562.

**[0018]** In the foregoing embodiments according to the invention were described. The invention is not limited to these embodiments. For example, in the scaling of the picture inset in the main picture it is possible to use interpolation also in the case of a common decoder. In that case, the picture to be reduced is decoded in a simple manner e.g. by selecting from the numbers produced by the DCT only the numbers representing the dc components of each block. Scaling is then performed using interpolation, followed by new encoding prior to the combination of the pictures. The extraction of different video signals from the transport stream may also be realized using a single demultiplexer instead of two or more. The invention idea may be applied in different ways within the scope defined by the independent claims.

## Claims

- 50 1. A method for inserting a moving secondary picture in a moving main picture, in which method the signals of said pictures are in an encoded digital format and the secondary picture is scaled down by reducing the number of macro blocks included in each individual frame of the picture and the scaled-down frame is inset as mini-picture in the frame of the main picture, characterized in that said inserting is performed prior to the decoding of the picture sig-

nals.

2. A method according to claim 1, characterized in that in order to inset the frame of the scaled-down secondary picture, or mini-picture, the code of the macro blocks in a certain area of the frame of the main picture are replaced by the code of the macro blocks of the frame of the mini-picture.

3. A method according to claim 1, characterized in that said reduction of the number of macro blocks in a frame is realized by leaving a selected number of macro blocks at regular intervals both in the horizontal and in the vertical dimension.

4. A method according to claim 1, characterized in that said reduction of the number of macro blocks in a frame is realized by compiling each new macro block from the blocks of at least two original macro blocks.

5. A method according to claim 1, characterized in that said reduction of the number of encoded macro blocks in a frame is realized by

- decoding the video signal to be scaled down by including in the decoding from each block at least the number that represents the dc component of the video signal,
- producing one new macro block by means of interpolation from a predetermined number of macro blocks produced, and
- encoding the signal produced using the same coding method as that used for the signal of the main picture.

6. A method according to claim 1, characterized in that there are at least two secondary pictures to be inset in the main picture.

7. A method according to claims 1 and 2, in which the code of the main picture and the code of the secondary picture are transmitted in fixed-form packets via the same transmission path as part of a transport stream, characterized in that the packets belonging to said pictures are extracted from the transport stream on the basis of identifiers in the headers of the packets and a coherent main picture code and coherent secondary picture code are generated and then combined.

8. An arrangement for insetting a moving secondary picture in a moving main picture, the signals of said pictures being in an encoded digital format and the arrangement comprising means (230) for scaling down each individual frame in the secondary picture and combining them with the frame of the main picture, characterized in that the arrangement also

comprises a decoder (210) for decoding the combined frame code into a video signal.

9. An arrangement according to claim 8, characterized in that the means for scaling down an individual frame and combining it with the frame of the main picture comprises a unit (231) for reducing the number of the macro blocks in a frame, a selector (232) for picture signals, and a timing unit (233) for placing the mini-picture at a certain location in the main picture.

10. An arrangement according to claims 8 and 9, characterized in that it also comprises means (541, 542) for extracting the encoded signal of the main picture from the transport stream comprised of data packets, and for extracting the encoded signal of the secondary picture from said transport stream.

11. A receiver adapted so as to combine the signals of at least two moving pictures in order to display said pictures simultaneously, characterized in that it comprises a decoder to decode the combined picture signal.

25

30

35

40

45

50

55

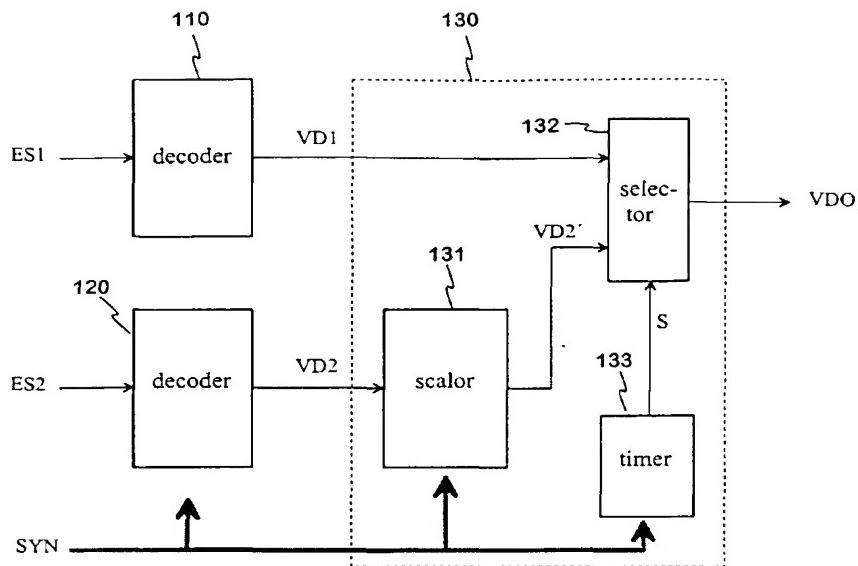


Fig. 3

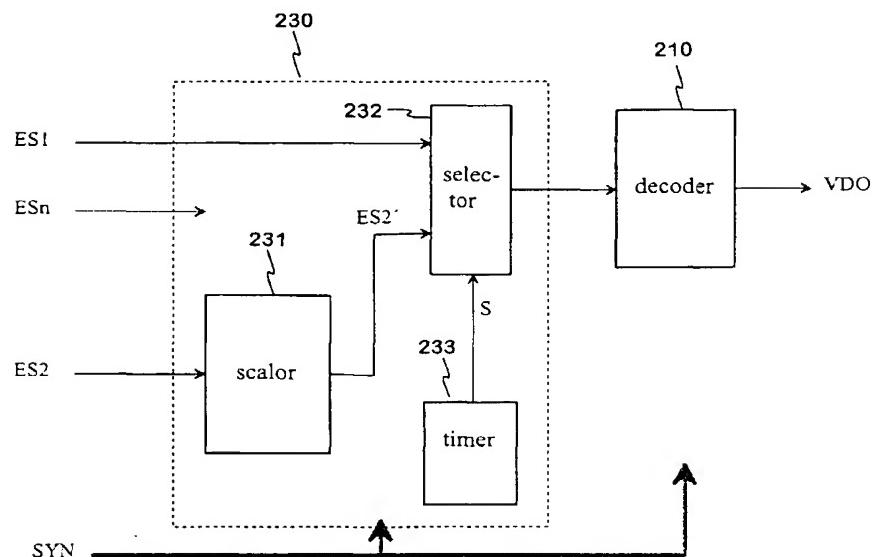
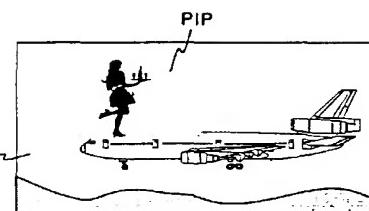


Fig. 2

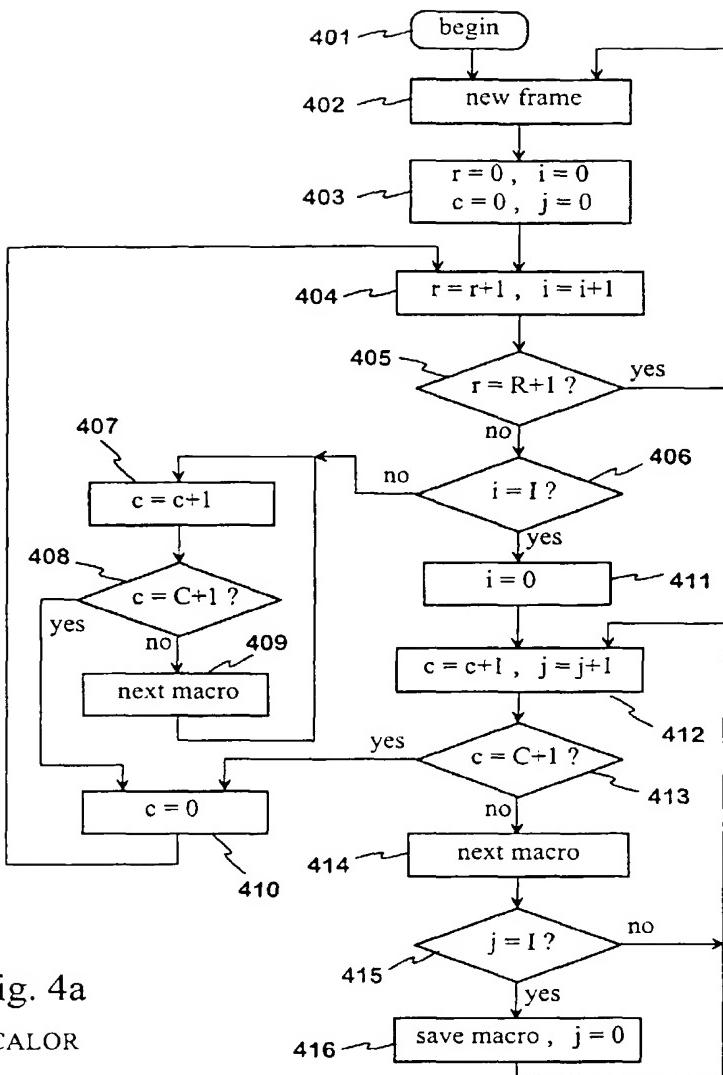


Fig. 4a

SCALOR

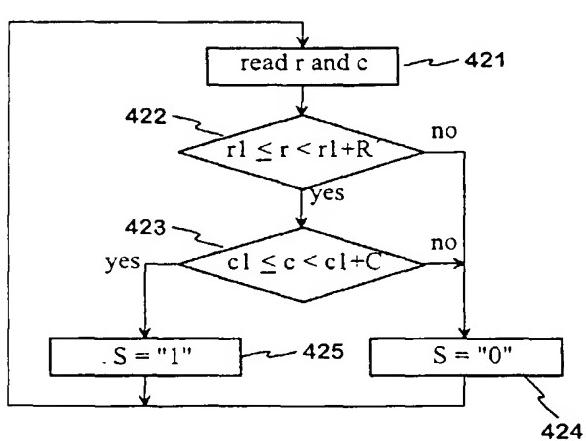


Fig. 4c

## TIMER

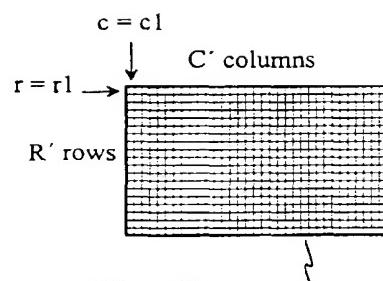


Fig. 4b

$$R' = I[R/I]$$

$$C' = I[C/I]$$

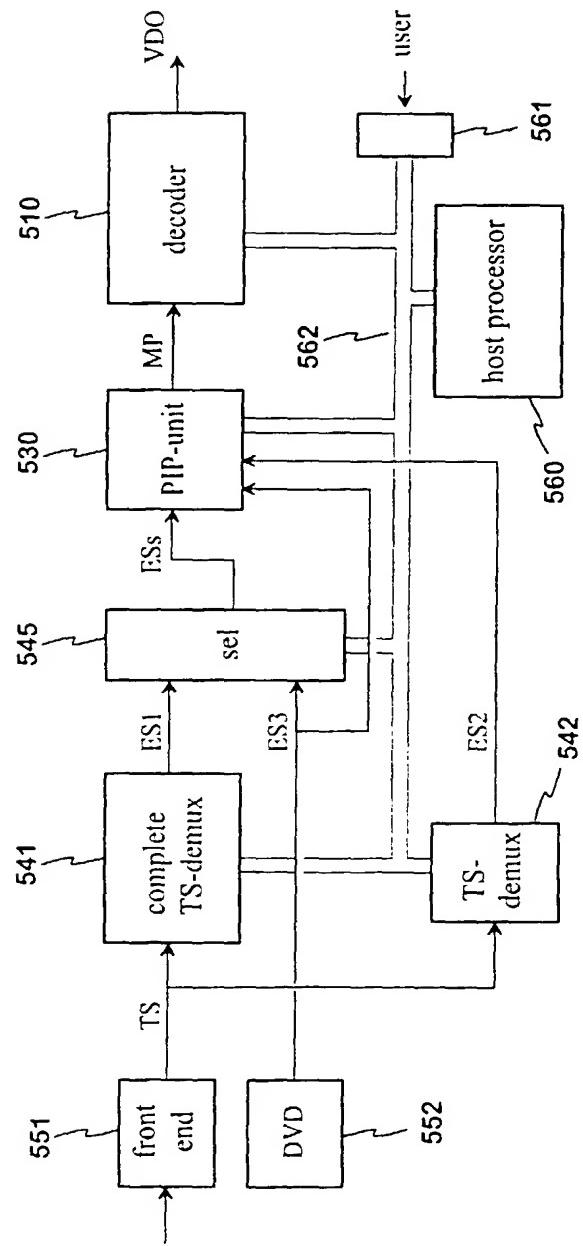


Fig. 5